DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

NOTES ON THE EXTENDED CLASS OF STEIN ESTIMATORS

TECHNICAL REPORT NO. 16

SUK-KI HAHN

JULY 1986

U. S. ARMY RESEARCH OFFICE
CONTRACT DAAG29-85-K-0239
THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

NOTES ON THE EXTENDED CLASS OF STEIN ESTIMATORS

Suk–ki Hahn
Korea Advanced Institute of Science and Technology
and
Stanford University

TECHNICAL REPORT NO. 16
July 1986

U. S. Army Research Office
Contract DAAG29-85-K-0239
Theodore W. Anderson, Project Director

Department of Statistics
Stanford University
Stanford, California

Approved for Public Release; Distribution Unlimited.

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

NOTES ON THE EXTENDED CLASS OF STEIN ESTIMATORS

Suk-ki Hahn

Department of Statistics
Stanford University
and
Korea Advanced Institute of Science and Technology

ABSTRACT

An estimator in the extended class of Stein estimators has two undesirable properties. For a small value of prior guess, it ignores the data. Moreover, for some cases its risk is not uniformly smaller than that of Stein estimator. We show that there exists a lower bound on $\tau(S)$ to guarantee a smaller risk, and the resulting estimator does not ignore the data.

1. Introduction and Motivation.

Consider a problem of estimating the mean vector θ of a $p \geq 3$ dimensional multivariate normal distribution on the basis of sample $X \sim N_p(\theta, I)$. Under the squared error loss, the maximum likelihood estimator $\delta_0(X) = X$ has risk $R(\theta, \delta) = p$ for every vector point of θ. James and Stein (1960) showed that the estimator

$$\delta_{JS}(X) = (1 - \frac{p-2}{S})X, \ \ S = \|X\|^2 = \sum_{i=1}^{p} X_i^2 \tag{1.1}$$

has risk $R(\theta, \delta_{JS}) < p$ for every θ. Even though it is uniformly better than the MLE we cannot use this estimator for the case of $S < p-2$. Our main purpose in this article is to show that there exists a class of estimators, containing a prior knowledge, whose members have a smaller risk than that of the Stein estimator. And also they give better protection against misspecification of the prior knowledge.

Sclove (1968) made an improvement using only the positive part of the Stein estimator. It is

$$\delta_{JS}^{+}(X) = (1 - \min\{1, \frac{p-2}{S}\})X, \ \ S > 0,$$

which satisfies the Baranchick (1970) conditions for minimax estimators. Efron and Morris (1973) interpreted the original estimation problem of the normal location parameter vector as an estimation problem for the hyper-parameter of a normal prior (distribution), $\theta \sim N_p(0, B^{-1}(1-B)I)$ with $B \in (0,1)$, under a "Relative Savings Loss" (RSL)

$$RSL(B, \delta) = \frac{R(B,\delta) - R(B,\delta^*)}{R(B,\delta_0) - R(B,\delta^*)} \tag{1.2}$$

where $R(B, \delta)$ is the expected risk of an estimator δ and $\delta^*(x) = (1-B)X$. They minimized $E_g RSL(B, \delta)$ over the Baranchick class of minimax estimators and then they derived the extended class of Stein estimators,

$$\delta_b^{c+}(X) = (1 - \min\{b, \frac{c(p-2)}{S}\})X, \ \ 1 \leq c \leq 2, \ \ 0 < b \leq 1, \tag{1.3}$$

where $E_g(\cdot)$ indicates expectation with respect to a hyper prior (distribution) $g(B)$ for the hyper-parameter B of the normal prior. We note that δ_{JS}^{+} is a member of the extended class of Stein estimators. The estimators δ_b^{c+} are not comparable with Stein estimators unless $b = 1 = c$. That is, for $b \neq 1$ and $c \neq 1$, there exists θ_1 and θ_2 ($\theta_1 \neq \theta_2$) such that $R(\theta_1, \delta_{JS}) < R(\theta_1, \delta_b^{c+})$ while $R(\theta_2, \delta_{JS}) > R(\theta_2, \delta_b^{c+})$. Only the positive part Stein estimator is uniformly better (in terms of risk) than δ_{JS}.

A natural question arises at this point. Suppose we have a prior knowledge (with a strong belief) which can be expressed by $N_p(0, b^{-1}(1-b)I)$ with known b,

and we want that the estimator we shall use has smaller risk than that of Stein estimator. Do we need to ignore the prior guess b and use δ_{JS}^{+} even though we believe that the linear estimator $(1-b)X$ is correct? This motivates us to seek for an improved estimator which allows the use of a prior knowledge.

Hence we have two criteria. The first one is that an improved estimator, containing a prior information, should have smaller risk than that of Stein estimator. Stein (1973) proposed a class of estimators which may have members dominating the positive part estimator. Efron and Morris (1976) gave a general class (a larger class than Alam's (1973)) of minimax estimators which allows $\tau(S)$ in $\delta = (1-\frac{2n}{S}\tau(S))X$ to decrease. Conditions for estimators with $\tau(S)$ strictly decreasing at some point with smaller risk than δ_{JS} has not yet been found. However, we restrict our attention to the Baranchik class of minimax estimators with $\lim_{S\to\infty}\tau(S)=1$, and we show that there exists a lower bound for $\tau(S)$ which leads to the better estimator. This is done in theorem 2 in section 2.

As the second criterion, an improved estimator must have good protection against a prior misspecification because (almost) always we have some useful information about the problem other than the sample. A Bayesian may hope for a posterior robustness over the all prior distributions while, based on sampling theory, the risk robustness (thus minimaxity) is desired. A plausible compromise between these two extremes may be a Bayes risk robustness over the all prior distributions. It is very difficult, at least for us, to work with a class of all prior distributions; thus we restrict the class of prior distributions to the class of normal distributions with zero mean vector and $B^{-1}(1-B)I$ covariance, indexed by the hyper-parameter B. We, therefore, adapt the relative savings loss (defined in (1.2) with $N_p(0, B^{-1}(1-B)I)$) which is a normalized version of a Bayes risk and is a function of B alone as a measure of protection against a wrong prior guess. Thus the estimator must have smaller RSL than δ_{JS} over the region of $B \in (0,1]$.

In summary, an improved estimator in the form of $\delta_H^b(X) = (1-\frac{2n}{S}\tau_H(b,S))X$ must satisfy the following conditions.

Condition 1. $\tau_H(b,S)$ is nondecreasing in $S>0$.

Condition 2. $0<\tau_H(b,S)\le \min\{1,(p-2)b/S\}, S>0 \;\; 1\ge b>0$.

Condition 3. $R(\theta,\delta_H^b)\le R(\theta,\delta_{JS})$ for all θ.

We note here that if Condition 3 is satisfied then $RSL(B,\delta_H^b)\le RSL(B,\delta_{JS})$ for all $B/b>0$ where b is a prior guess for B. To choose the best one among estimators in

$$D = \{\text{estimators which satisfy above 3 conditions}\},$$

we minimize $RSL(b,\delta)$ over the class D. We find that estimators with, for $2n = p-2$,

$$(1.4) \quad 2n\tau_H(b,S) = \begin{cases} bSI_{(0,2n/b)}(S) + 2nI_{(2n/b,\infty)}(S), & n/(n+1) \le b \le 1, \\ \gamma_n(S)I_{(0,S_b)}(S) + bSI_{(S_b,2n/b)}(S) + 2nI_{(2n/b,\infty)}(S), & 0 < b \le n/(n+1) \end{cases}$$

satisfy the above three conditions and minimize $RSL(b,\delta)$ over the estimators in D. This is done in Section 3. We note here that S_b is the solution of $\gamma_n(s) = bs$ for $b \in (0, n/(n+1)]$ and that $\gamma_n(s)$ is defined by

$$(1.5) \qquad \gamma_n(s) = s\int_0^1 t^n \exp(-ts/2)dt \Big/ \int_0^1 t^{n-1}\exp(-ts/2)dt.$$

Its properties are given in the appendix.

2. Main Result.

Estimators that contain a prior knowledge and have smaller risk than that of Stein estimator are desired. For this purpose we start with an estimator with absolutely continuous $\tau(s)$ in order to get a lower bound on τ.

Theorem 1. (Efron and Morris (1976)). Suppose τ is absolutely continuous with derivative τ'. If the risk $R(\theta,\delta)$ is finite and if the expectation of each term in (2.1) exists, then a unique unbiased estimator of $R(\theta,\delta)$ based on the sample, S, exists and is given by

$$(2.1) \qquad \hat{R}(\theta,\delta) = p - (p-2)[\frac{p-2}{S}\tau(S)(2-\tau(S)) + 4\tau'(S)].$$

This theorem implies that nondecreasing condition of τ is not necessary for an estimator to be minimax, but no convenient substitution for this condition has been found. We, therefore, keep the nondecreasing condition in Baranchick's (1970) theorem. The Baranchick class of minimax estimators is too large since it contains some estimators that are not better than the Stein estimator. One way to guarantee that the estimator we will use is better than the Stein estimator is to make it satisfy the conditions in the following theorem.

Theorem 2. If the absolutely continuous function $\tau(s)$ with derivative $\tau'(s)$ satisfies the conditions for any $S > 0$ and $p = 2(n+1) \ge 3$,

i) $\tau(\cdot)$ is non-decreasing,

ii) $\gamma_n(S)/2n \le \tau(S) \le \min\{2n/S, 1\}$,

then $R(\theta,\delta) \le R(\theta,\delta_{JS})$ for all θ. The equality holds when $|\theta|^2 = 0$ and $\tau(S) = \gamma_n(S)/2n$.

Proof. From Theorem 1,

$$
\begin{aligned}
R(\theta,\delta) &= p - 2nE_\theta\{\frac{2n}{S}\tau(S)(2-\tau(S)) + 4\tau'(S)\} \\
&= \{p - 2nE_\theta\frac{2n}{S}\} + 2nE_\theta\{\frac{2n}{S}(1-\tau(S))^2 - 4\tau'(S)\} \\
&= R(\theta,\delta_{JS}) + 2nE_\theta\{\frac{2n}{S}(1-\tau(S))^2 - 4\tau'(S)\}.
\end{aligned}
\tag{2.2}
$$

Thus it is enough to show that

$$
E_\theta[\frac{2n}{S}(1-\tau(S))^2 - 4\tau'(S)] \le 0 \quad \text{for all} \quad \theta. \tag{2.3}
$$

Since S is distributed as chi-square with p degrees of freedom and noncentrality parameter $\lambda = \|\theta\|^2/2$, (2.3) can be expressed as

$$
\sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!}E_{p+2k}[\frac{2n}{S}(1-\tau(S))^2 - 4\tau'(S)] = \sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!}R_k(\tau), \tag{2.4}
$$

where $E_{p+2k}(\cdot)$ indicates the expectation with respect to the central chi-square distribution with $p+2k$ degrees of freedom. For the chi-square distribution, it can be shown by integration by parts that

$$
E_m(S - m\mu)h(S) = 2\mu E_m S h'(S) \tag{2.5}
$$

for $S \sim \mu\cdot\chi^2_m$ and $h(\cdot)$ such that all expectations in (2.5) exist. Using this equality, $R_k(\tau)$ in (2.4) can be rewritten as

$$
R_k(\tau) = \frac{n}{n+k}E_{2(n+k)}\{1-\tau(S)\}^2 - \frac{1}{n+k}E_{2(n+k)}\{[S-2(n+k)]\tau(S)\}
$$

since

$$
\begin{aligned}
4E_{m+2}(\tau'(S)) &= 4E_m(\frac{S}{m}\tau'(S)) \\
&= \frac{2}{m}E_m\{[S-m]\tau(S)\}
\end{aligned}
$$

with $m = 2(n+k)$. Further, by integration by parts,

$$
\begin{aligned}
R_k(\tau) &= \frac{1}{2(n+k)}\{2n + 4kE_{2(n+k)}\tau(S) + 2nE_{2(n+k)}(\tau(S))^2 - 4(n+k)E_{p+2k}\tau(S)\} \\
&\propto n(c-1)^2 + \int\{2(n+k)F_{p+2k}(s) - (2k+2n\tau(s))F_{2(n+k)}(s)\}d\tau,
\end{aligned}
\tag{2.6}
$$

where $c = \lim_{s\to\infty}\tau(s) = 1$ from condition i) and ii), thus the first term vanishes. The integrand in the second term can be rewritten as

$$
F_{2(n+k)}(s)\cdot[2(n+k)\frac{\gamma_{n+k}(s)}{2(n+k)} - 2k - 2n\tau(s)], \tag{2.7}
$$

where $\gamma_{n+k}(s)/2(n+k) = F_{2(n+k+1)}(s)/F_{2(n+k)}(s)$ with $\gamma_n(\cdot)$ defined in (1.5). (See the appendix for the several properties of $\gamma_n(\cdot)$ function.) We also know that $2-\gamma_{n+1}(s)+\gamma_n(s) \geq 0$ for every $n > 0$ and $s > 0$. Thus the expression in braces in (2.6) can be rewritten as

$$\begin{aligned}
&\gamma_{n+k}(s) - 2k - \gamma_n(s) + \gamma_n(s) - 2n\tau(s) \\
&= \gamma_n(s) - 2n\tau(s) - \sum_{j=1}^{k}(2 + \gamma_{n+j-1}(s) - \gamma_{n+j}(s)) \\
&\leq \gamma_n(s) - 2n\tau(s) \\
&\leq 0 \quad \text{from condition (ii).}
\end{aligned}$$

Combining this and condition (i) of nondecreasing $\tau(s)$, the expression for $R_k(\tau)$ in (2.6) gives $R_k(\tau) \leq 0$ for every integer $k \geq 0$; thus the theorem is proved. We note that, when $\|\theta\|^2 = 0$ (i.e. $\lambda = 0$), if we choose $\tau(s) = \gamma_n(s)/2n$ then the quantity in brackets in (2.7) becomes

$$\gamma_n(s) - 2n\tau(s) = \gamma_n(s) - \gamma_n(s) = 0$$

and the Poisson random variable has nonzero weight only when $k = 0$. This proves the assertion of the equality. Q.E.D.

3. Construction.

If $\tau(s)$ is not absolutely continuous then the expression (2.1) does not exist. However, the lower bound $\gamma_n(s)/2n$ for $\tau(s)$ is very useful searching for a better estimator. Define

$$\tau(s) = \begin{cases} \sum_{i=1}^{3} \tau_i(s) I_{(S_{i-1},S_i)}(s), & \text{if } S_i \text{ exists for } i = 1,2, \\ \tau_1(s), & \text{otherwise,} \end{cases} \tag{2.8}$$

where $\tau_i(S_i) = \tau_{i+1}(S_i)$ for $i = 1,2, S_0 = 0, S_3 = \infty$ and $\tau_i(s)$ are absolutely continuous in $(S_{i-1}, S_i]$. Then, with some conditions on $\tau_i(s)$, such $\tau(s)$ gives a smaller risk than that of the Stein estimator.

Theorem 3. An estimator $\delta(X) = (1-2n\tau(S)/S)X$ with $\tau(s)$ defined in (2.8) has smaller risk than that of the Stein estimator if, for any value of $s, \gamma_n(s)/2n \leq \tau(s) \leq \min\{1, s/2n\}$, where $p = 2(n+1)$ and $\tau(s)$ is nondecreasing.

Proof. If S_i does not exist for $i = 1,2$ then, since $\gamma_n(s)/2n \leq \tau_1(s) = \tau(s) \leq \min\{1, S/2n\}$ for any $S > 0$ and if $\tau_1(s)$ is absolutely continuous, this theorem is proved from Theorem 1. The risk difference can be expressed as

$$R(\theta,\delta) - R(\theta,\delta_{JS}) = \sum_{j=0}^{\infty} \frac{\lambda^j \exp(-\lambda)}{j!} R_j(\tau)$$

where

$$R_j(\tau) = E_{p+2j}[S(1-\frac{2n}{s}\tau(s))^2 - 4j(1-\frac{m}{s}\tau(s)) + 2j] - p + 4n^2/2(n+j)$$
$$\propto n(1-2E_{p+2j}\tau(s) + E_{2(n+j)}[\tau(s)]^2) - 2j(E_{p+2j}\tau(s) - E_{2(n+j)}\tau(s)). \tag{2.9}$$

We know that if F_0, F_1 are two cumulative distribution functions on the real line such that $F_1(x) \le F_0(x)$ for all x, then $E_0h(X) \le E_1h(X)$ for any nondecreasing function $h(\cdot)$. Thus, $E_{p+2j}\tau(s) - E_{p+2j-2}\tau(s) \ge 0$ since our $\tau(s)$ is nondecreasing and $F_{p+2j} \le F_{p+2j-2}$. It is, therefore, enough to show that the first term in (2.9) is nonpositive. For $k = 3$,

$$R_{j1} = 1 - 2E_{p+2j}\tau(s) + E_{p+2j-2}(\tau(s))^2$$
$$= 1 - \sum_{i=1}^{3}\{2\int_{S_{i-1}}^{S_i} \tau_i(s)dF_{p+2j}(s) - \int_{S_{i-1}}^{S_i} (\tau_i(s))^2 dF_{p+2j-2}(s)\}. \tag{2.10}$$

We note here that, for all $S \in (S_2, \infty)$, $\tau_3(s) = 1$. In each interval $(S_{i-1}, S_i]$, (2.10) can be rewritten by integration by parts, as

$$\begin{aligned}
R_{j1} &= 1 - 2\sum_{i=1}^{3}\{F_2(S_i)\tau_i(S_i) - F_2(S_{i-1})\tau_i(S_{i-1}) - \int_{S_{i-1}}^{S_i} F_2(s)d\tau_i(s)\} \\
&+ \sum_{i=1}^{3}\{F_1(S_i)[\tau_i(S_i)]^2 - F_1(S_{i-1})[\tau_i(S_{i-1})]^2 - 2\int_{S_{i-1}}^{S_i} \tau_i(s)F_1(s)d\tau_i(s)\} \\
&\propto \sum_{i=1}^{3}\int_{S_{i-1}}^{S_i} (F_2(s) - F_1(s)\tau_i(s))d\tau_i(s) \\
&= \sum_{i=1}^{3}\int_{S_{i-1}}^{S-i} F_1(s)[\frac{\gamma_{n+k}(s)}{2(n+k)} - \tau_i(s)]d\tau_i(s),
\end{aligned}$$

where $F_1(s) = F_{p+2j-2}(s)$ and $F_2(s) = F_{p+2j}(s)$. From the property (viii) of the $\gamma_n(\cdot)$ function (see the appendix), $\gamma_{n+k}(S)/2(n+k) \le \gamma_n(S)/2n \le \tau_i(s)$ for all $S \in (S_{i-1}, S_i], i = 1, 2, 3$. This proves the result for $k = 3$. For $k = 2$, it can be shown easily by putting $S_2 = \infty$ and $\tau_2(s) = 1$ for all $s > S_1$. Q.E.D.

Any estimator defined in Theorem 3 has uniformly smaller risk than that of the Stein estimator. Thus, from the definition, so does the RSL. To choose the best among them, we minimize $RSL(b, \delta)$ under the restrictions that $\gamma_n(S)/2n \le \tau(S) \le \min\{1, S/2n\}$ and nondecreasing $\tau(s)$. Let S_b be a solution of $\gamma_n(s)/s = b$.

Theorem 4. The estimator in the class of estimators defined in Theorem 3 which minimizes $RSL(b, \delta)$ is given by $\delta_H^b(X) = (1 - \frac{2n}{S}\tau_H(b, S))X$, where

$$\tau_H(b, S) = \begin{cases} \gamma_n(S)/2nI_{(0,S_b)}(S) + bS/2nI_{(S_b,2n/b)}(S) + I_{(2n/b,\infty)(S)}, & \text{if } S_b \text{ exists,} \\ bS/2nI_{(0,2n/b)}(S) + I_{(2n/b,\infty)}(S), & \text{otherwise.} \end{cases}$$

Proof. We note here that the condition that S_b exists can be replaced by that of $0 < b \le (p-2)/p$ since for any $s > 0, \gamma_n(s)/s \le (p-2)/p$. It can be shown that

$$RSL(b,\delta) = E_{p+2}\{(\frac{2n}{bS}\tau(S) - 1)^2|b\}$$
$$= E_{p+2}\{\{\frac{2n}{bS}(\tau(S) - \frac{Sb}{2n})\}^2|b\} \propto E_p\{(\tau(S) - \frac{Sb}{2n})^2|b\}$$

is minimized at $\tau(S) = bS/2n$. Imposing the restriction that $\gamma_n(S) \le 2n\tau(S) \le \min\{S, 1/2n\}$, that we get $RSL(b, \delta_H^b) = \min RSL(b,\delta)$, where the minimization is over the etimators defined in this theorem.

4. Evaluation and Comments.

1. The lower bound $\gamma_n(S)/S$ of shrinkage has an interesting property when S approaches to zero. In the Bayesian framework with normal prior $N_p(0, B^{-1}(1-B)I)$, the marginal distribution of X is $N_p(0, B^{-1}I)$. When we have $X = 0$ (thus $S = 0$), then the (empirical) Bayesian estimator of B will be

$$E^{\{X=0\}}B = \frac{\int_0^1 B \cdot B^{n+1}\exp(-BS/2)g(B)dB}{\int_0^1 B^{n+1}\exp(-BS/2A)g(B)dB}\Bigg|_{S=0}$$
$$= \int_0^1 B^{n+2}g(B)dB \Big/ \int_0^1 B^{n+1}g(B)dB \le 1,$$

where the equality holds if and only if the hyper prior $g(B)$ is concentrated at $B = 1$. It depends only on the prior information. When we do not have any information about B and we use $g(B) \propto B^{-2}$ (a limiting case of Strawderman (1971)), then $E^{\{X=0\}}B = \frac{n}{n+1} = \lim_{S\to 0}\frac{\gamma_n(S)}{S}$. We note again $p - 2 = 2n$.

2. Another interesting property is that it makes it easy to put some prior information about B, say $b \in (0,1]$, into the estimation procedure. One example using the lower bound is an estimator with $\tau_H(b,S)$ in (1.5). It has a good property which an estimator in the extended class of Stein estimators defined in (1.4) with $c = 1$ does not possess. Berger (1982) gave it an intuitive justification as being the Bayes estimator (based on $\theta \sim N_p(0, b^{-1}(1-b)I)$) when the prior guess b is supported by the data (small s), and being a Stein estimator otherwise. The null hypothesis of $B = b$ is rejected if the data turn out to be small (near zero), and we can infer that B is bigger than b, but $\hat{B}_b^+(S)$ remains in b. This undesirable property of $\hat{B}_b^+(S)$ becomes severe when b approaches zero. That is $\lim_{b\to 0}\hat{B}_b^+(S) = 0$ no matter what the data are. When we have almost zero prior knowledge (almost uniform distribution on θ) then the extended class of Stein estimators becomes MLE; thus the risk remains p for any value of θ. But $B_H(b,S) = (p-2)\tau_H(b,S)/S$ with $\tau_H(b,S)$ in (1.4) gives $\gamma_n(S)/S$ when b

approaches zero and we know that its risk $R(\theta, \delta_H)$ is smaller (uniformly) than that of the Stein estimator; thus the effect of the lower bound $\gamma_n(S)/S$ is great.

3. Another merit of $B_H(b, S)$ is that it gives a very stable protection against misspecification of prior information. This can be explained in terms of relative savings loss. Berger (1982) expressed the RSL of estimators in the extended class of Stein estimators as a function of B/b.

Theorem 5. (Berger (1982)). Define $2n = p - 2$ and $\lambda = B/b$, where B and b are true and prior value of hyper parameter of normal distribution $N(0, B^{-1}(1-B)I)$. Then

$$RSL(B, \delta_b^+) = (1 - \lambda^{-1})^2 + A_1(\lambda) \cdot [1 - F_{p+2}(2n\lambda)] + A_2(\lambda) \cdot \beta \cdot f_{p+2}(\lambda|2n),$$

where

$A_1(\lambda) = 2/p - (1 - \lambda^{-1})^2$,

$A_2(\lambda) = 1/p - 1/(p-2)\lambda$,

$f_{p+2}(\lambda|2n) = 2n\chi^2_{p+2}$,

$F_{p+2}(\cdot)$ is the cdf of the chi-square distribution with $p + 2$ degrees of freedom.

The values of RSL when $p = 4$ for various λ are in Table 1. It shows that if $B > 3b$ then $RSL(B, \delta_b^+) > 0.5 = RSL(B, \delta_{JS})$ while, for any value of $B/b, RSL(B, \delta_H^b) < 0.5$.

TABLE 1. $RSL(B, \delta_b^+), p = 4$

λ	RSL	λ	RSL
0	0.5	1.5	0.42
.1	0.469	3	0.505
.3	0.423	4	0.584
.5	0.393	5	0.648
.7	0.376	10	0.810
0.9	0.370	100	0.980
1.0	0.368	∞	1.0

4. An analogue of the generalized prior distribution on θ, which Berger (1980) suggested using is

$$h_p(\theta) = \int_0 (\frac{c}{B} - 1)^{-p/2} \exp[-\frac{B \cdot \|\theta\|^2}{2(c-B)}\} B^{-2} dB.$$

It is a heavy tail prior, the tail chosen to yield robustness (on the prior). This leads to the Bayes estimator $(1 - \frac{\gamma_n(s/c)}{s})X$. The constant c can be interpreted as

$$c = B\{1 + \text{ prior guess for common variance in } N_p(0, \tau I)\}.$$

But it can be less than B; thus the variance term $c/B - 1$ can be negative. To avoid this difficulty, the range of integration is modified by $B \in (0, c)$ not $(0, 1)$. Then

$$\begin{aligned} h_p(\theta) &= \int_0^c (c/B - 1)^{-p/2} \exp\{-\frac{B \cdot \|\theta\|^2}{2(c - B)}\} B^{-2} dB \\ &\propto \int_0^1 (B^{-1} - 1)^{-p/2} \exp\{-\frac{B \cdot \|\theta\|^2}{2(1 - B)}\} B^{-2} dB \end{aligned}$$

and this leads to the Bayes estimator $(1-\gamma_n(s)/s)X$. We note here that the shrinkage estimator due to the robust prior distribution is the lower bound of $\tau(s)$.

5. δ_H^b has an empirical Bayes property. The fact that $\lim_{n\to\infty} 2n/S = B$ with probability one is known. Thus with the expression $\gamma_n(S)/2n = 1 - \{\sum_{i=0}^{\infty}(s/2n)^i \frac{\Gamma(n+i)n^i}{\Gamma(n+i+1)}\}^{-1}$, it can be shown that $\lim_{n\to\infty} \gamma_n(S)/2n = 1$ with probability one and these imply that both bounds $\gamma_n(S)/S$ and $2n/S$ approach to the true value of B with probability one for the case of large p (thus large n). So does $B_H(b, S)$. This implies that, for large $p, \delta_H^b(X)$ is very close to the optimal linear estimator $(1 - B)X$.

APPENDIX

1. Properties of $\gamma_n(s)$.

For any $n > 0$ and $s > 0$,

i) $0 < \gamma_n(s) < 2n$.

ii) $\gamma_n(s)$ is increasing in s.

iii) $\gamma_n(s)$ is increasing in s.

iii) $\lim_{s \to 0} \gamma_n(s)/s = n/(n+1)$.

iv) $\gamma_n(s)/s$ is decreasing in s.

v) $\lim_{n \to \infty} \gamma_n(s) = s$.

vi) $0 < \gamma_{n+1}(s) - \gamma_n(s) < 2$.

vii) $\gamma_n(s)/2n = F_{2(n+1)}(s)/F_{2n}(s)$ where $F_{2n}(s)$ is the cdf of the chi-square distribution with $2n$ degrees of freedom.

viii) $\gamma_{n+1}(s)/\gamma_n(s)$ is decreasing in n.

ix) $\gamma_{n+1}(s) - \gamma_n(s)$ is increasing in s.

Proof. From (i) to (vi), the proof is in Berger (1980). For the part (vii),

$$\begin{aligned}\frac{\gamma_n(s)}{2n} &= \frac{s}{2n} \frac{\int_0^1 t^n \exp(-ts/2)dt}{\int_0^1 t^{n-1} \exp(-ts/2)dt} \\ &= \frac{\Gamma(n)2^n}{\Gamma(n+1)2^n} \frac{\int_0^s x^n \exp(-x/2)dx}{\int_0^s x^{n-1} \exp(-x/2)dx} = \frac{F_{2(n+1)}(s)}{F_{2n}(s)}.\end{aligned}$$

For part (viii), suppose there exist some s, say S_0, such that

$$\gamma_{n+2}(S_0) - \gamma_{n+1}(S_0) \geq \gamma_{n+1}(S_0) - \gamma_n(S_0)$$

and this is true for any $n > 0$. We know that $\gamma_{n+1}(S_0) - \gamma_n(S_0) > 0$ for any $n > 0$ from part (vi). But part (v) gives $\lim_{n \to \infty} (\gamma_{n+2}(S_0) - \gamma_{n+1}(S_0)) = 0$. Therefore, there exists no such S_0, and thus the assertion follows. With this and the fact

$$\begin{aligned}\frac{\partial}{\partial s}(\gamma_{n+1}(s) - \gamma_n(s)) &= \frac{1}{2s}\{\gamma_{n+1}(s)A_{n+1}(s) - \gamma_n(s)A_n(s)\} \\ &\propto [\{\gamma_{n+1}(s) - \gamma_n(s)\}\{2 + \gamma_n(s)\} - \gamma_{n+1}(s)\{\gamma_{n+2}(s) - \gamma_{n+1}(s)\}] \\ &> (\gamma_{n+1}(s) - \gamma_n(s)) - (\gamma_{n+2}(s) - \gamma_{n+1}(s)) \\ &> 0 \quad \text{from part (viii)},\end{aligned}$$

part (ix) is clear. We note here that $A_n(s) = 2 - \gamma_{n+1}(s) + \gamma_n(s)$.

REFERENCES

1. Alam, K. (1973). "A Family of Admissible Minimax Estimators of the Mean of a Multivariate Normal Distribution," *Annals of Statistics*, 1, 517–525.
2. Baranchik, A. J. (1970). "A Family of Minimax Estimators of the Mean of a Multivariate Normal Distribution," *Annals of Mathematical Statistics*, 41, 642–645.
3. Berger, J. (1980). "A Robust Generalized Bayes Estimator and Confidence Region for a Multivariate Normal Mean," *Annals of Statistics*, 8, 716–761.
4. Berger, J. (1982). "Bayesian Robustness and Stein Effect," *Journal of the American Statistical Association*, 77, 358–368.
5. Efron, B. and Morris, C. (1973). "Stein's Estimation Rule and Its Competitors: An Empirical Bayes Approach," *Journal of the American Statistical Association*, 68, 117–130.
6. Efron, B. and Morris, C. (1976). "Families of Minimax Estimators of the Mean of a Multivariate Normal Distribution," *Annals of Statistics*, 4, 11–21.
7. Sclove, S. L. (1968). "Improved Estimators for Coefficients in Linear Regression," *Journal of the American Statistical Association*, 63, 596–606.
8. Strawderman, W. (1971). "Proper Bayes Minimax Estimators of Multivariate Normal Mean," *Annals of Mathematical Statistics*, 42, 385–388.
9. James, W. and Stein, C. (1960). "Estimation with Quadratic Loss," *Proceedings of the Fourth Berkeley Symposium* in *Mathematical Statistics and Probability*, 1, 311–319, University of California Press.

Technical Reports
U.S. Army Research Office
Contracts DAAG29-82-K-015 and DAAG29-82-K-0239

1. "Maximum Likelihood Estimators and Likelihood Ratio Criteria for Multivariate Elliptically Contoured Distributions," T. W. Anderson and Kai–Tai Fang, September 1982.

2. "A Review and Some Extensions of Takemura's Generalizations of Cochran's Theorem," George P.H. Styan, September 1982.

3. "Some Further Applications of Finite Difference Operators," Kai–Tai Fang, September 1982.

4. "Rank Additivity and Matrix Polynomials," George P.H. Styan and Akimichi Takemura, September 1982.

5. "The Problem of Selecting a Given Number of Representative Points in a Normal Population and a Generalized Mills' Ratio," Kai–Tai Fang and Shu–Dong He, October 1982.

6. "Tensor Analysis of ANOVA Decomposition," Akimichi Takemura, November 1982.

7. "A Statistical Approach to Zonal Polynomials," Akimichi Takemura, January 1983.

8. "Orthogonally Invariant Minimax Estimator of the Covariance Matrix of a Multivariate Normal Population," Akimichi Takemura, April 1983.

9. "An Orthogonally Invariant Minimax Estimator of the Covariance Matrix of a Multivariate Normal Population," Akimichi Takemura, April 1983.

10. "Relationships Among Classes of Spherical Matrix Distributions," Kai–Tai Fang and Han–Feng Chen, April 1984.

11. "A Generalization of Autocorrelation and Partial Autocorrelation Functions Useful for Identification of ARMA(p,q) Processes," Akimichi Takemura, May 1984.

12. "Methods and Applications of Time Series Analysis Part II: Linear Stochastic Models," T. W. Anderson and N. D. Singpurwalla, October 1984.

13. "Why Do Noninvertible Estimated Moving Averages Occur?" T. W. Anderson and Akimichi Takemura, November 1984.

14. "Invariant Tests and Likelihood Ratio Tests for Multivariate Elliptically Contoured Distributions," Huang Hsu, May 1985.

15. "Statistical Inferences in Cross–lagged Panel Studies," Lawrence S. Mayer, November 1985.

16. "Notes on the Extended Class of Stein Estimators," Suk–ki Hahn, July 1986.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 16	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) NOTES ON THE EXTENDED CLASS OF STEIN ESTIMATORS		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER --
7. AUTHOR(s) Suk-ki Hahn		8. CONTRACT OR GRANT NUMBER(s) DAAG29-85-K-0239
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, Sequoia Hall Stanford University Stanford, CA 94305-4065		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P-19065-M
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE July 1986
		13. NUMBER OF PAGES 12pp.
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An estimator in the extended class of Stein estimators has two undesirable properties. For a small value of prior guess, it ignores the data. Moreover, for some cases its risk is not uniformly smaller than that of Stein estimator. We show that there exists a lower bound on $\tau(S)$ to guarantee a smaller risk, and the resulting estimator does not ignore the data.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)